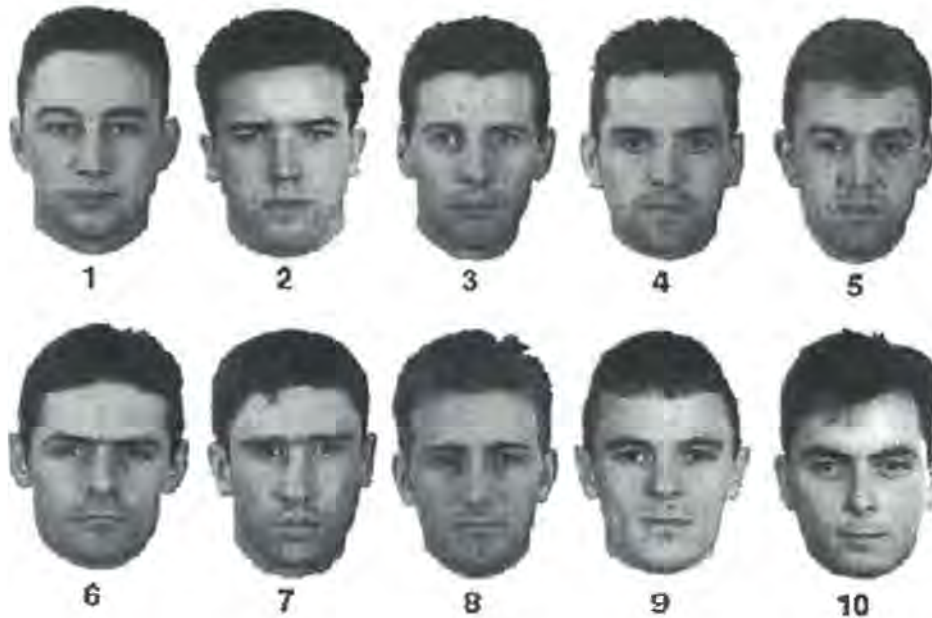The following is an example of a 10-person array used with the face above. Participants were reminded that the target face may or may not be in the array.

Performance on what seems like a relatively simple and straight-forward test of face recognition proved surprisingly poor: only 59.5% of participants correctly identified the face when it was present in the array (another 9.5% erroneously picked someone else). When the target-person was not in the array, 23.5% erroneously picked someone from the array. Because half the arrays contained the target and half did not, only 64% of all selections were correct (59.5/(59.5+9.5+23.5)). Note that one might argue that the 23.5% of witnesses shown a target-absent array who nonetheless erroneously identified someone are essentially making calculated guesses. Furthermore, because participants were randomly assigned to target-absent and target-present arrays, one could further argue that a comparable 23.5% of the choosers in the target-present arrays were similarly making calculated guesses (9.5% miscalculated and picked a non-target member of the array and another 14% guessed correctly – thus, nearly 1 in 4 of the correct identifications of the target may be little more than calculated guesses … choices of the person who most-resembles the target.

Immediate Recognition of Two Faces from Memory. In a second study Megreya & Burton (2006) simply presented 2 faces sequentially (on the theory that people encounter others sequentially) and then tested recognition (for one or the other person) in similar 10-person photoarrays (with the same reminder that one of the persons may or may not be in the array. This time the correct identification rate fell to 31.2%, the false identification rate in target-present arrays was 25.8% and the false identification rate in target-absent arrays was 39.2%--just 32% of all selections were correct. As per the argument above, it would appear that nearly 40% of witnesses were making calculated guesses.

Immediate Recognition of Faces—Matching from Pictures. In a third study the same researchers allowed the participants to view the targets while they made their selections from the photoarrays (this is termed a matching task and does not involve any memory—the target and the identification-array are viewed simultaneously). In this matching task performance was also surprisingly poor: the target was selected just 70% of the time from target-present arrays (with a 12.3% false identification rate) and in target-absent arrays, the false identification rate was 33.6% (just 60% of selections were correct). Note that the rate of calculated guessing was still very high even though the decisions were not based on memory.

Immediate Recognition of Faces—Matching Live Persons to Pictures from Memory. Megreya & Burton (2008) pushed the testing further: in one study participants were shown a live person for 30 seconds and were then tested, from memory, on a 10-person photoarray—when the target was in the array he was identified by 70% of the participants and a non-target person was identified by 10% of participants. When the target was not in the array, 20.5% misidentified someone else (70% of all positive choices were correct— with a calculated-guessing rate of 20.5%).

Immediate Recognition of Faces—Matching Live Persons to Pictures. When participants were shown the live person and the photoarray together (a matching task rather than a memory test), the target was identified by 66.9% of the participants and a non-target person was identified by 15% of participants. When the target was not in the array, 37.8% misidentified someone else (thus, 55.9% of all positive choices were correct and for some reason the opportunity to view the target nearly doubled the calculated-guessing rate). Finally, when the target was shown together with just a single photograph and the target was in the photograph he was identified by 84.4% of the participants and when the target was not in the photograph, 15.6% misidentified that person. In this instance the calculated-guessing rate is at least 15.6% but we cannot know how many participants guessed the target.

The difficulties inherent in recognizing faces is underscored by the following set of pictures from Jenkins, et al. (2011) who pose the simple question—how many different people are shown in the photographs. The answer is given on the signature page below.

Studies of Real Witness Performance Using Police Archives Show High Rates of Identification Errors

Similarly low accuracy rates from actual eyewitness identifications emerge from studies of actual witnesses. In these studies it is not known whether the suspect is the actual perpetrator, but it is still possible to gauge the rate of inaccurate identifications of foils—the known-innocents placed in arrays along with the suspects. The best data come from studies in the United Kingdom—the results from nearly 17,000 actual eyewitnesses are shown in the table below.

UK Archival Studies –9-person arrays [row percentages]				
	Witnesses/ Arrays	Suspect ID	Foil ID	No ID
Slater	843/302	36%	22%	42%
Wright & McDaid	1,561/616	39%	20%	41%
Valentine, et al. (VIPER)	584/295	41%	21%	39%
Pike,et al.	8,800/xx	49%	-	-
W Yorkshire-traditional	1,635/xx	35%	-	-
W Yorkshire-VIPER	940/xx	39%	-	-
Wright & (VIPER)	134/134	58%	21%	21%
Horry & (VIPER)	1,359/xx	40%	26%	34%
Memon & (VIPER)	1,044/xx	44%	42%	15%
Ave [complete-wtd]	16,897/xx	40%	26%	34%

As shown the table, Slater (1994) examined identification attempts by 843 British witnesses who viewed 302 suspects. Slater found that foils were mistakenly identified by 190 witnesses (22.5 percent). Similarly, Wright and McDaid (1996) examined identification attempts of 1,561 British witnesses who viewed 616 suspects in live lineups and found that 611 witnesses (39.1 percent) picked the suspect, 310 witnesses picked a known-innocent foil (19.9 percent), and 640 (41 percent) made no identification. In all, these studies report results on nearly 17,000 identification attempts. Among the studies where complete data are available, nearly *40% of positive identifications were identifications of an innocent foil* (26% foils / (26% foils + 40% suspects) = 39.4%). While these studies are informative with respect to the rates at which witnesses appear willing to make choices/guesses from arrays (we know, authoritatively that foil IDs are incorrect) archival studies cannot tell us whether suspect IDs or lineup rejections are correct – for an extensive consideration of the weaknesses inherent in archival studies, see Horry, et al. (2014).

Witness Are Frequently Guessing

Witness error rates are likely to be significantly higher than these numbers immediately indicate due to another problem: it is likely that some of those arrays did not contain the perpetrator. If 75% of those arrays contain the perpetrator and 25% have innocent suspects and innocent and guilty suspects are selected equally often by witnesses, it turns out that less than half of witnesses (45%) who make a selection pick a guilty person (.40 suspect selection rate x .75 guilty rate = 30% guilty picks) and more than half pick an innocent person (.40 suspect selection rate x .25 not-guilty rate = 10% innocent suspects + 26% innocent foils selections = 36% innocent selections (more than one-fourth of whom are innocent suspects who are now in trouble).

But performance is likely affected by another factor that is not obvious in the overall numbers. It is likely that witnesses in these studies were confronted with nine-person arrays, the British standard. This would mean that the average foil drew guesses from 3.3% of witnesses (26/8=3.3)—if these were fair arrays (a dubious proposition in light of evidence of lineup bias) we would expect that another 3.3% of witnesses simply guessed the suspect in the same way that other witnesses guessed innocent foils. If the arrays were biased in a typical manner (see discussion below), it is more likely that between 6% and 10% of witnesses guessed the suspect (some of whom are innocent)—if so, then as few as 40%-10%=30% of witnesses might have selected the suspect based on a sound memory. In sum, a combination of poor memory, willingness to guess, a modest percentage of innocent suspects and small arrays conspire to create a situation in which a substantial portion of suspect identifications are innocent suspects (around one-quarter in the examples above).

Overall, 34% of witnesses did not identify anyone in the array—these non-identifications reflect one of two types of error—either the witnesses "missed" a guilty suspect or the police offered witnesses an array which did not contain the perpetrator. Some reliable witnesses would correctly reject arrays which do not contain the guilty party—other witnesses may (luckily) reject the array because their memory is poor and they are not inclined to make a guess (Penrod, 2003).

Behrman and Davey (2001) examined actual Sacramento, California police records for 271 cases involving 349 crimes (the vast majority armed robberies). They examined 289 photographic lineups and 58 live lineups. They found that approximately 48% of the witnesses who observed a photographic lineup identified the suspect as the perpetrator. Unfortunately, they do not know how many of those were identifications of the actual perpetrator as opposed to mistaken identifications and, because police records were incomplete, Behrman and Davey also could not determine how many additional identifications were made of known-innocent foils—the numbers from the UK studies suggest that 20-25% of witnesses would have identified foils. In live lineups (which did report mistaken identifications of foils—probably due to the presence of defense counsel), Behrman and Davey found that 50% of the witnesses identified the suspect (an unknown percentage of those identifications were erroneous identifications of innocent suspects), the false identification of foils rate was 24% and 26% of the witnesses rejected the lineup. Here, *one in three positive identifications was clearly wrong,* and, it is likely that between 10% and 15% of the 50% suspect IDS were merely guesses (using the logic applied above with respect to UK arrays—with the US arrays producing higher rates of suspect guesses because the arrays have 6 rather than 9 members) and either the witness or the police made a mistake with respect to the cases in which no ID was made.. In short, errors were rampant.

Probably the best data on witness performance in American lineups come from recent field studies. Wells, Steblay & Dysart (2011) studied 497 witnesses who attempted identifications from photoarrays. Overall 26% of witnesses picked a suspect and 15% picked a foil – as in the studies from the UK a substantial portion of witnesses making a selection clearly made an error (15/(15+26) = 37%. Because these were 6 person arrays, the 15% who chose a foil were selecting foils at an average rate of 3% (15%/5) – and if these arrays were biased to the extent of arrays studied by others (see below), there is a good chance that 6%-9% of the 26% of witnesses who picked a suspect similarly guessed the suspect (that is, between 23% and 35% of suspect selections might be characterized as "calculated guesses")!

Wixted and colleagues (2016) report the results from 717 six-person photospreads conducted by the Houston Police Department. These resulted in IDs of suspects by 32% of witnesses, filler identifications by 29% and no identifications by 39%--thus, nearly 48% of positive identifications (29/(29+32)) were clearly erroneous and a substantial percentage (perhaps 12% to 18%) of the 32% of witnesses selecting the suspect were similarly guessing. Wixted, et al. estimate that arrays were target-present (contained the guilty perpetrator) in 35% of these arrays and 65% contained an innocent suspect.

Further evidence of the unreliability of identifications comes from studies by Sagana and her colleagues (2013, 2016). In their 2013 study 71 pedestrians were engaged in conversation by two experimenters

pretending to be tourists. After a short interval, the pedestrians were asked to identify each of the experimenters from separate six-person photarrays. They were then shown their selections and asked to explain the reasons for their decisions. However, for recognition decisions, the lineup member chosen by the pedestrian was swapped with an unidentified member of the array. More than 40% of pedestrians did not notice the swap. The 2016 study examined blindness for identifications from target-present (TP) and target-absent (TA) arrays. Eighty pedestrians were exposed to a staged theft and asked to identify the thief and the victim from separate six-person arrays. In this study identification decisions concerning the thief were manipulated such that participants' selections were swapped with a previously unidentified lineup member and lineup rejections were converted into identifications.

In short, in both experimental studies and real police cases, one-third or more of witnesses who make a positive identification are wrong and a high proportion of witnesses are guessing—those who guess a foil reveal their errors but someone else—police, prosecutors, judges or jurors--has to detect the errors made by witnesses who guess suspects.

The "Pleading Effect" Elevates the Portion of Trial Defendants Who Have Been Misidentified

Though these analyses imply that a significant portion of identified suspects are innocent, the rates of mistaken identification that appear for trial are likely to much higher than the rates of mistaken identifications. As Wells, Memon and I (2006) note, archival research indicates people charged with crimes plead guilty in a high percentage of cases—(Liptak [2011] reports the rate was 97% in federal courts in 2010 and 94% in state courts in 2006). In contrast Garrett (2008) reports that just 4% of the defendants in the 200 DNA-exonerated cases he studied entered a guilty plea. What impact might this have on the rate of innocent suspects in court? An archival analysis (Kellstrand, 2006) of cases with both DNA and eyewitness identification evidence in the San Diego County District Attorney's Office, reported that the individuals suspected of crimes were innocent in only 5% of cases. If we assume that just 5% of suspects identified are innocent and 95% are guilty these numbers imply that in state courts, 6% of the 95% (5.7%) of the guilty go to trial, whereas 96% of the 5% (4.8%) of the innocent suspects go to trial. Consequently, at the trial level, nearly half (45.7%) of the defendants (4.8% of the 10.5% going to trial) will be instances of mistaken identification (at the federal level the mix is 2.9% guilty + 4.8% innocent go to trial and 62% are innocent). Charman and Wells (2006) termed this the "pleading effect" and it illustrates how the mistaken-identification rate can be much, much higher at the trial level than at the lineup level.

VIII. Witnessing Factors that Influence the Accuracy of Identifications

Weapon Focus

As noted at the beginning of Section VII above, the record indicates both intruders were carrying weapons.

Weapon focus refers to the attention witnesses give to a perpetrator's weapon during a crime. It is expected that the attention the witness focuses on a weapon will reduce their ability to later recognize the perpetrator. Researchers have assessed eyewitness accuracy in an attempt to assess the effects of weapon-focus effects on memory; the relevant studies were reviewed in a meta-analysis by Steblay (1992). The meta-analysis included 19 studies with a total sample of 2,082 participants. The weapon-focus effect on identifications was statistically significant and reflected a modest impairment overall—however, the effect was larger in target-absent lineups and when memory was generally impaired.

A good illustration of the sort of impairment that can be obtained under realistic conditions comes from a study by Maass & Kohnken (1989). In this study 86 non-psychology research participants were approached by an experimenter who was either holding a syringe or a pen and either did or did not threaten to administer an injection. The confederate held, at participants' eye level, a transparent plastic syringe with a blue label and needle protection cap that was partially filled with a yellow liquid. In the no-threat condition, the

confederate (holding either a syringe or a pen) apologized for disturbing the experiment and explained she had to pick up a drug used in a different experiment. She then put the pen or syringe on a tray, took a medicine bottles and left the room. Participants were asked to identify the target from a seven-person, target-absent array. As shown in the table, exposure to the syringe greatly decreased lineup recognition (while enhancing the accuracy of recall for hand cues). The results also show that threat/stress similarly increased error rates.

Table 2. Percentages of False Alarms as a Function of Syringe and Threat of Injection

	Syringe absent	Syringe present	Total
No threat	33.3% ($n = 21$)	63.6% ($n = 22$)	48.8%
Threat	57.1% ($n = 21$)	68.2% ($n = 22$)	62.8%
Total	45.2% ($n = 42$)	65.9% ($n = 44$)	

In a recent dissertation study by one of our students at John Jay College (by an active duty Connecticut police chief) DeCarlo (2010) used a videotaped robbery and found that in a no weapon condition, witnesses were able to correctly identify the target 78% of the time. When a weapon was implied by the perpetrator waving his hands around in his pocket, accuracy dropped to 55% and when a weapon was actually shown, accuracy dropped to 33%--those latter two conditions and impairment are comparable to my 1987 study. When the perpetrator was absent from the lineup the correct rejection was 89% for the no weapon condition, but when a weapon was implied, accuracy dropped to 76% and when a weapon was actually shown, correct rejections dropped to 65%. A number of studies have demonstrated that a variety of surprising objects (including a whole, raw chicken) can draw attention away from the perpetrator and that novelty, rather than threat, is likely to be the critical ingredient in the effect—indeed, the study by Maass & Kohnken alluded to above, demonstrated either novelty/attention-grabbing or threat can independently produce impairment.

Research by Mitchell, Livosky, and Mather (1998); Pickel (1998, 1999); and Shaw and Skolnick (1999) indicates that any surprising object can draw attention away from the perpetrator and that novelty, rather than threat, may be the critical ingredient in the effect. Researchers have tried to detect weapon-focus effects in field studies, and although the results are conflicting (likely because the methods used are fraught with problems detailed by Pickel (2007)), Tollestrup, Turtle, and Yuille (1994) examined the effect of weapon focus on the rate of suspect identification and obtained data consistent with laboratory findings.

A new meta-analysis (Fawcett, et al., 2013) confirms these conclusions using data from 47 comparisons and further notes that the size of the effects was "unaffected by whether the event occurred in a laboratory, simulation, or real-world environment."

As emphasized earlier, research shows that jurors overestimate the accuracy of identifications, jurors fail to distinguish accurate from inaccurate eyewitnesses, jurors tend to ignore viewing conditions that are known to predict identification accuracy and instead based their decisions in part on eyewitness memory for peripheral details and witness confidence--both of which tend to be poor predictors of identification accuracy. In Desmarais & Read's (2010) meta-analysis of 23 surveys assessing 4,669 lay respondents' knowledge of eyewitness issues, they reported that just 47% of American respondents understood that weapon focus would impair identification performance (though as detailed above that does not mean that this understanding would be reflected in their assessments of trial evidence). Clearly, there is substantial disagreement among laypersons about the impact that the presence of a weapon can have on identification performance and this is not the kind of disagreement that can really be resolved through logic or argument— as with all the factors discussed below, the nature of the relationship between weapon presence and

identification accuracy is an empirical, scientific question—eyewitness expert testimony would provide jurors with scientifically-grounded guidance on this issue and, as detailed above, increase juror sensitivity to weapon focus.

Tracking the Identity of Multiple Individuals

As noted at the beginning of Section VII above, the record indicates there were two intruders.

It has been demonstrated in a variety of contexts that witnesses have significant difficulty keeping track of the identity of multiple individuals. Just as a weapon can divide the attention of a witness, multiple perpetrators can divide their attention and identification accuracy and accuracy of descriptions have been shown to decline as the number of perpetrators increases (Clifford and Hollin, 1981; Fahsing, Ask, and Granhag, 2004; Megreya, et al 2006; Shepherd, 1983). This impairment effect was clearly illustrated in the dissertation research by our John Jay student DeCarlo (2010) who showed his witnesses a videotaped robbery involving one or two perpetrators. When the perpetrator was present in the lineup, the single perpetrator condition produced a correct identification rate of 55% versus 33% in the multiple perpetrator condition. In the multiple perpetrator condition, 82% of witnesses correctly rejected perpetrator-absent arrays versus 92% in the single perpetrator condition. More recently, Megreya & Bindeman (2012) compared the performance of witnesses seeing one or two perpetrators. Their data showed that when the perpetrator was present in an array fewer correct identifications (29%) and more incorrect decisions (71%) were made for the double-perpetrator condition compared to the single-perpetrator condition (54% and 46%, respectively). Over a third of witnesses made misidentifications from perpetrator-absent arrays. The authors note: "the double-perpetrator disadvantage is not simply due to a difficulty in separating the facial characteristics of two people that are encountered simultaneously, which therefore cannot be encoded and stored accurately (see Palermo & Rhodes, 2002). Instead, we suggest that this effect might reflect the need to divide attention between two concurrent culprits, which limits the depth or detail of the stored descriptions that can be formed for these persons (see Fahsing et al., 2004)." p.448

Neither Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues nor Benton et al. (2006) provides insight into lay understanding of the effects of multiple perpetrators on identification accuracy. Because both multiple perpetrators and weapon focus appear to involve distracted attention, perhaps the results reported above with respect to understanding of weapon focus (47% believing weapon presence impairs performance) provides some guidance about the likely disagreement among laypersons about the impact of this factor.

Assessing the Probative Value of Identifications, Non-Identifications and Multiple Identifications

As noted at the beginning of Section VII above, the record indicates there are several identifications of Mr. Nolan and one non-identification. How, if at all, do these decisions cumulate?

Multiple-witness cases are fairly common: archival studies report that multiple-witness cases comprised 29% of 662 cases studied by Yuille & Tollestrup (1992), 38% of 76 cases studied by Tollestrup et al. (1994) and 58% of lineups involving 661 suspects in a study by Wright and McDaid (1996). The Innocence Project (http://www.innocenceproject.org/docs/Eyewitness_ID_Report.pdf) reports that 38% of the DNA exoneration cases involving eyewitness identification had more than one witness making the same misidentification and in 13% of cases there were 3 misidentifications.

In 1974 Robert Buckhout published an article in *Scientific American* which contained the following photographs of two different individuals who had been mistakenly identified and arrested for crimes actually committed by the person in the middle picture.

I have personally been involved in a case where three witnesses misidentified an innocent person who was convicted and cleared while in prison when the real perpetrator was apprehended after committing similar crimes (see Deckinga picture below) and a case with two misidentifications where the innocent person was cleared while awaiting sentencing (see Stevens picture below). A classic case in the literature involved seven misidentifications of a priest as a purse-snatcher. The real culprit, who wore a hat while committing the crimes, was identified during the priest's trial—after the identifying witnesses had already testified (see picture below—from Ellison & Buckhout, 1981). Although the actual culprit was 14 years younger than the priest, if one covers their heads, one sees a similarity between the two men.

In these instances of misidentification (and most other I am familiar with) the innocent person and the actual perpetrator bore a resemblance to one another and similarity of appearance can clearly be one of the ingredients contributing to misidentifications.

Comparison
Shaun Deckinga vs. Jerome Clapper

Deckinga
Birth: 3-?-?
Age: ??
Height: ??
Weight: ??
Eyes: Brown
Hair: ??

Clapper
Birth: ?-?-??
Age: 40
Height: ??
Weight: 180 lb.
Eyes: Blue
Hair: Brown

Work by Clark & Wells (2008) provides us with the most extensive analysis of multiple "identifications" and their potential weaknesses. Clark and Wells assume independence of eyewitnesses—which is to say that one eyewitness's decision does not influence other eyewitness's decisions. They assume equivalence among witnesses—e.g., that witnesses had equally good opportunities to view the perpetrator, received the same instructions when making identification attempts, viewed the same lineup, made their identification attempts at equal time periods after a crime, etc.

Clark & Wells note that they: "consider TP (Target Present) and TA (Target Absent) lineups together to evaluate diagnosticity... diagnosticity refers to the probability that the suspect is guilty given a particular response, whereas accuracy refers to the probability of a particular response given that the suspect is guilty or

innocent. From the standpoint of the trier of fact, diagnosticity is the key question, which can only be addressed by considering TP and TA lineups together."

Clark & Wells use standard Bayesian analysis to examine a number of scenarios observed in the research literature. For these analyses they assume that half of lineups are TP and half are TA ()—this means that the "prior probability" the suspect is guilty is .5--and they illustrate the calculation or the "posterior probability" the suspect is guilty following various witness responses using the examples in their Table 1 below.
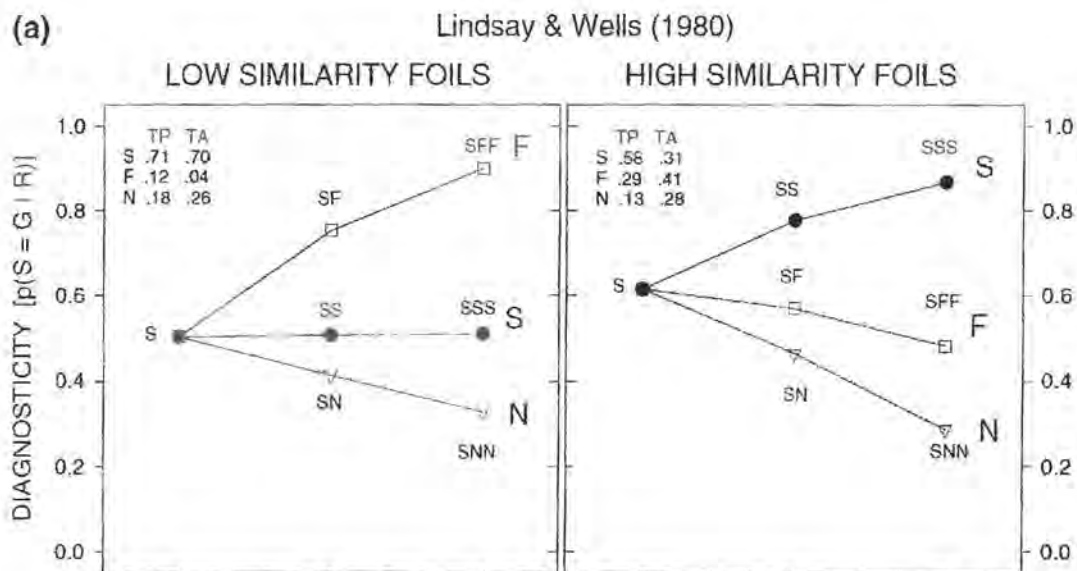
Table 1 Example calculations for single- and multiple-witness diagnosticities

	TP	TA	Single witness	Two witnesses
Suspect	.5	.1	.5/.5 + .1 = .833	S + S (.5)(.5)/(.5)(.5) + (.1)(.1) = .962
Foil	.2	.3	.2/.2 + .3 = .400	S + F (.5)(.2)/(.5)(.2) + (.1)(.3) = .769
No ID	.3	.6	.3/.3 + .6 = .333	S + N (.5)(.3)/(.5)(.3) + (.1)(.6) = .714

Note: Two-witness calculations are shown for two suspect identifications (S + S), one suspect and one foil identification (S + F), and one suspect and one nonidentification (S + N). TP = target-present and TA = target-absent
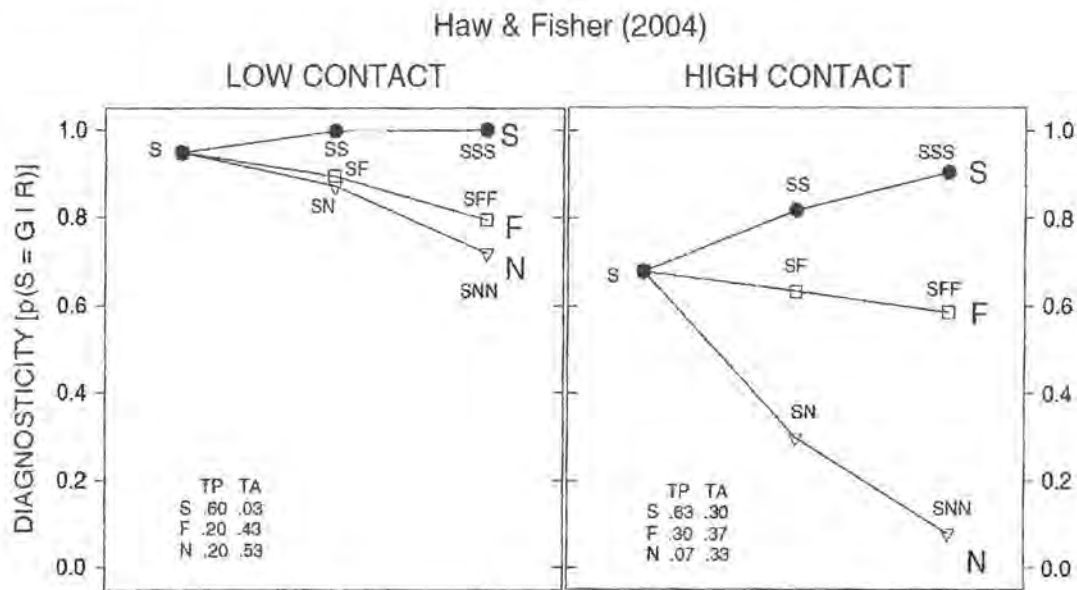
Clark and Wells examine a number of scenarios (using observed levels of performance from studies noted in the bottom left of each panel) using various combinations of suspect (S), filler (F) and no-choice (N) conditions. Note that suspects may be guilty or innocent depending on whether they are selected from TP or TA arrays. One can see that under varying conditions the diagnosticity of IDs is reduced or enhanced. These include variations the fairness of photoarrays and identification procedures. Perhaps the most important point, for present purposes, to emerge from these analyses is that multiple suspect identifications (SS and SSS) can, under some conditions, add nothing or very little to the probability the suspect is actually guilty.

Low similarity foils (which make the innocent suspect stand out) yield lower diagnosticity for suspect identifications.

(a) Lindsay & Wells (1980)

Note that with the low similarity foils used in the example study *suspect identifications had no diagnostic value* (because the guilty and innocent suspects were equally likely to be chosen)—indeed, only foil

identifications had diagnostic value with respect to the suspects guilt (only because foil choices were more frequent in target-present arrays—an outcome that would be useless to police.

Haw & Fisher (2004)

The Probative Value of Non-Identifications

As noted at the beginning of Section VII above, the record indicates

Though naïve logic may suggest that multiple eyewitness identifications are stronger evidence of guilt than single identifications, it turns out that logic can and does break down under a number of circumstances (as evidenced by the DNA exonerations and my examples). For the past 30 years the eyewitness research community has recognized that non-identifications can be substantial evidence that a suspect is not guilty. As early as 1980 Wells and Lindsey questioned the criminal justice system's policy of treating eyewitness identifications of suspects as highly probative while treating non-identifications (both in the form of no-choices and foil choices) as non-probative. Wells and Lindsey used a Bayesian model of information to demonstrate that if an identification of a suspect increases the probability the suspect is the perpetrator, then a nonidentification must reduce that probability.

They illustrated these results using data from a study by Loftus (1976) in which (a) when the perpetrator was present in the lineup, 84% of witnesses chose him, 4% of witnesses made nonidentifications, and 12% chose a foil; (b) when the perpetrator was absent, 60% chose an innocent suspect 24% made nonidentifications and 16% chose foils. Wells & Lindsay compared the information gained (increase in the probability the suspect is guilty following a suspect identification or decrease in the probability the suspect is guilty following a nonidentification (both "none of the above" plus foil identifications). As shown in the figure below, nonidentifications were more probative of innocence than suspect identifications were of guilt.
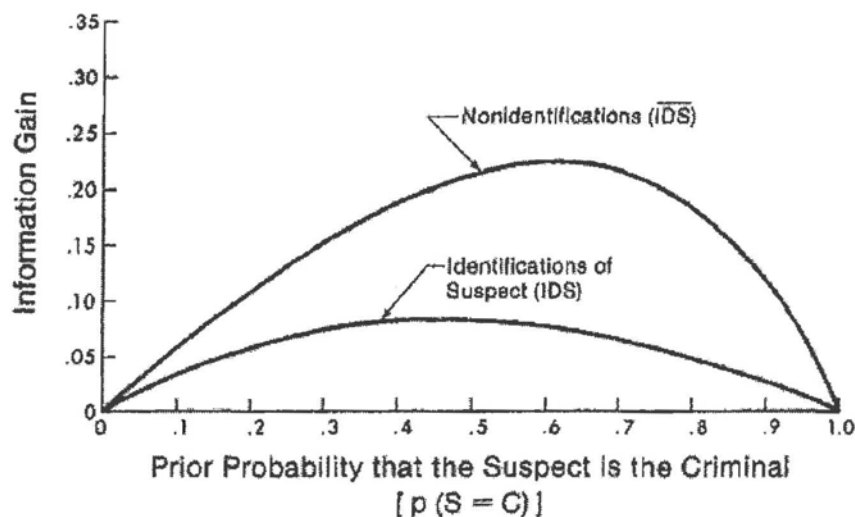
Figure 1. Information gain from identifications and nonidentifications regarding the probability that the suspect is the criminal as a function of the prior probability that the suspect is the criminal (using Loftus's, 1976, data).

Wells & Lindsay then compared the probative value of "none of the above" and foil identification responses and as shown in the figure below, found that "none of the above" responses were much more probative than foil identifications (which were roughly comparable to the probative value of suspect identifications as shown in the prior figure.
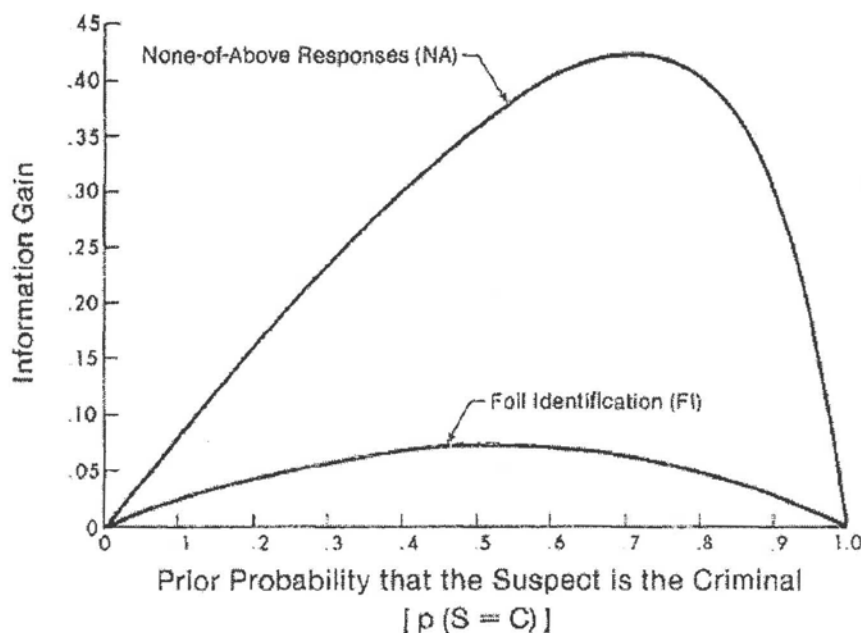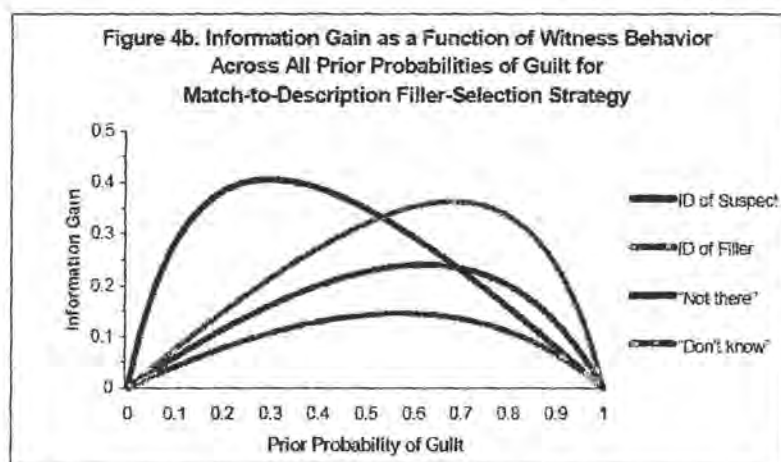

Figure 2. Information gain from none of-above choices and choices of foils regarding the probability that the suspect is the criminal as a function of the prior probability that the suspect is the criminal (using Loftus's, 1976, data).

Wells and Olson (2002) carried the Bayesian analysis further by examining the information gain (changes in the probability the suspect is the perpetrator) as a function of the prior probability the suspect is guilty (based on evidence prior to the identification results). They combined data from three different studies of sequential and simultaneous lineups and the figure below provides a good summary of their findings: filler identifications, non-identifications and "don't know" responses all have probative value with respect to the

suspect's innocence and sometimes filler identifications and non-identifications have more probative value (with respect to innocence) than positive identifications of the suspect have with respect to guilt.

Figure 4b. Information Gain as a Function of Witness Behavior Across All Prior Probabilities of Guilt for Match-to-Description Filler-Selection Strategy

Information Gain (y-axis: 0 to 0.5)
Prior Probability of Guilt (x-axis: 0 to 1)

Legend:
- ID of Suspect
- ID of Filler
- "Not there"
- "Don't know"

Clark, Howell and Davey (2008) carried these types of analyses further in a meta-analysis of 94 comparisons between target-present and target-absent lineups. Their analyses showed that suspect identifications were more diagnostic regarding a suspect's guilt or innocence than other responses, but that non-identifications were also diagnostic of the suspect's innocence. They note that their analyses indicate that a suspect identification is less informative if the lineup is biased and that the probative value of a suspect identification is undermined if suggestive lineup instructions increase the willingness of witnesses to make an identification. They note: "Nonidentifications also are straightforward. They are diagnostic of the suspect's innocence. We reiterate the point made by Wells and Lindsay (1980) that nonidentifications are not merely "failures" to identify the suspect, but rather carry important information whose value should not be overlooked" (p. 211). They further note that foil identifications appear to be diagnostic of innocence when foils are selected based on their match to the description of the perpetrator given by a witness, but not when foils are selected on their match to the suspect

Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues did not include non-identifications as a factor, nor did others such as Benton et al. (2006). As noted above, research on other factors certainly suggests there will be disagreement among laypersons about the impact that these factors have on identification performance—eyewitness expert testimony would provide jurors with scientifically-grounded guidance on this issue.

Stress

As noted at the beginning of Section VII above, the record indicates that various of the victims were threatened by weapons, assaulted, injured, and/or tied up.

Despite the importance of knowledge in this area, one cannot simulate violent crimes or pose a threat to the well-being of naive experimental subjects. Researchers have therefore resorted to a variety of manipulations including the use of violent versus nonviolent videotaped crimes. Increased violence in videotaped reenactments of crimes has been shown to lead to decrements in both identification accuracy and eyewitness recall (Clifford & Hollin, 1981; Clifford & Scott, 1978; Johnson & Scott, 1976; Sanders & Warnick, 1980).

Deffenbacher (1983, 1991) pointed to the "Yerkes-Dodson Law" when explaining the effects of arousal on identification. Stress or arousal demonstrates an inverted U-shaped relationship with identification accuracy. Low levels of arousal, such as when waking up, produce low attentiveness; moderate levels of arousal, such

as that felt by an athlete preparing to compete, serve to heighten perceptual and attentiveness skills; and, higher levels, such as that felt by an individual under extreme danger or duress, debilitates perceptual skills.

Deffenbacher, other colleagues, and I recently (2004) published a meta-analysis of stress effect studies. The meta-analysis was conducted on 27 tests of the effects of heightened stress on identification accuracy. This sample included work published between 1974 and 1997, with a total of 1727 participants involved in relevant tests of the stress-- the mean proportions correct for TP lineups under high and low stress conditions were .39 and .59, respectively (with false alarm rates of 34% and 19% respectively). The effect of stress was larger for target-present than for target-absent lineups—that is, stress particularly reduced correct identification rates. The effect was twice as large for eyewitness-identification studies that simulated eyewitness conditions (e.g., staged crimes) than for studies that induced stress in other ways.

These effects are confirmed and extended in a study by Morgan et al. (2004) that was published at the time of our paper and therefore not included in our meta-analysis. Morgan et al. examined the eyewitness capabilities of more than 500 active-duty military personnel enrolled in a survival-school program (see Table 3). After 12 hours of confinement in a mock prisoner-of-war camp, participants experienced both a high stress interrogation with real physical confrontation and a low-stress interrogation without physical confrontation. Both interrogations were 40 minutes long; they were conducted by different persons. A day after release from the camp, and having recovered from food and sleep deprivation, the participants viewed a 15-person live lineup, a 16-person photo spread, or a sequential presentation of photos of up to 16 persons. Regardless of the testing method, as the Table below shows, memory accuracy for the high-stress interrogator was much lower overall than for the low-stress interrogator.

	High Stress	Low Stress
HITS [Target-Present]		
Live line-up method	27%	62%
Photo-spread method	36%	76%
Sequential photo method	49%	75%
FALSE ALARMS [Target-absent]		
Live line-up method	45%	50%
Photo-spread method	48%	61%
Sequential photo method	0%	0%

Another more recent confirmation of findings comes from a study by Valentine & Mesout (2008) conducted in the "Horror Labyrinth" of the London Dungeon tourist attraction. The eyewitness study was conducted using self-report measures of anxiety previously validated against heart-rate changes. Tourists were recruited to participate in the study and they encountered a "scary person" while slowly walking around the labyrinth for approximately 7 minutes. About 45 minutes later, after they completed their tour the purpose of the experiment was explained to them and they were tested to see if they could identify the scary person from a nine-person photo-array. Complete data were obtained from 56 participants who were divided into two groups reflecting the 50% with the highest reported anxiety scores and the 50% with the lowest scores. As shown in the table below, 75% of the witnesses experiencing lower anxiety were able to identify the scary person/culprit but only 18% of those with higher anxiety. Over 50% of the higher-anxiety group incorrectly identified/guessed one of the eight "innocent" foils (which logically suggests that two of the five higher-anxiety witnesses also simply guessed the culprit).

Table 3. The outcome of identification attempts from a target-present nine-person photograph lineup as a function of state anxiety

	Low state anxiety	High state anxiety	Total
Culprit ID	21	5	26
Foil ID	6	15	21
No ID	1	8	9
Total	28	28	56

Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues did not include stress as a factor, but Benton et al. (2006) reported that 68% of American respondents believed that stress could impair identification performance. This means there is still substantial disagreement among laypersons about the impact that stress can have on identification performance—eyewitness expert testimony can provide jurors with scientifically-grounded guidance on this issue and, as detailed above, increase juror sensitivity to the effects of stress.
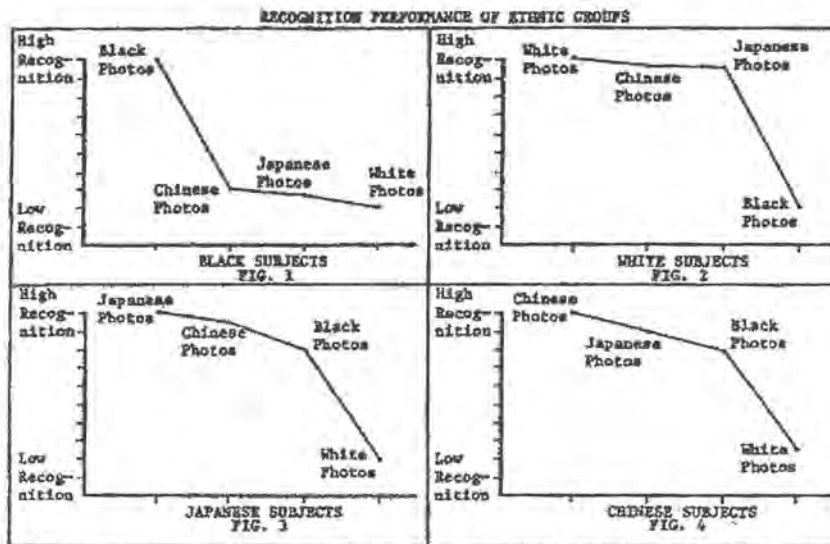
Cross-Race Identification

As noted at the beginning of Section VII above, the record indicates that witnesses gave varying descriptions of the perpetrators as Hispanic or White. The witnesses are Black and Hispanic – this raises the possibility of cross-race impairment in identification performance.

Research on cross-race identification impairment began forty years ago and has included various mixes of Caucasian, Asian, Hispanic, Black, and middle-eastern witnesses. For example, Platz and Hosch (1988) conducted a field study examining the identification performance of Mexican American, Black, and White convenience store workers in identifying customers who had interacted with them earlier in the day. A Mexican-American, Black, or White "customer" went into the store and made a fairly involved purchase or asked for directions from the clerk. Two hours later a pair of students posing as law office interns asked the clerk for help in identifying the individual using a series of lineup photos. Platz and Hosch found a significant cross-race impairment for all three of the racial groups: clerks of each group were better recognizing customers of their own race than customers from either of the other two races.

Luce (1974) tested the face recognition ability (for White, Black, Chinese and Japanese faces) of 65 Chinese and 60 Japanese S's who attended a California state university. As emphasized by Luce, all had been raised exclusively in California, lived in the multi-ethnic city of San Francisco, attended a multi-ethnic university, attended school with all ethnic groups, and reported specific friendships among these ethnic groups. Both groups displayed some impairment with Black faces as compared to their own ethnic group (Figures 3 and 4 below).

Luce concluded (based also on impaired cross-race performance by Whites and Blacks): "With respect to other-race recognition, the current study reveals that despite lifelong association with a given ethnic group, there is no assurance that its members will be readily recognized. The study, therefore, offers tentative support for the "they all look alike" notion on the part of some ethnic groups towards other ethnic groups. Finally, implications of these findings for witness identification are clear. Reliable recognition performance cannot be assured whenever a member of one ethnic group is called upon to attempt to recognize a member of a different ethnic group. Perhaps the clain (sic) "Yes that's the man who held me up", must be looked upon with considerable skepticism when the alleged culprit is of a different ethnic group than is his identifier." (p. 41)
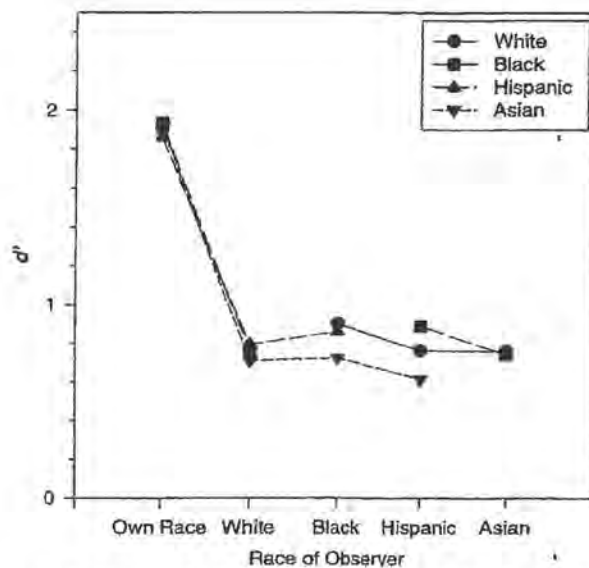
The performance of several racial groups (including Whites, Blacks, Latinos, and Asians) in recognizing both White and Black faces was examined by Teitelbaum and Geiselman (1997). Asian participants showed significantly lower hit scores on the Black faces than Black participants--as shown in this table (d' is a measure of "discriminability"--how effectively participants can identify faces)

Mean *d'* Score as a Function of Race of Observer, Race of To-Be-Remembered Faces, and Induced Mood

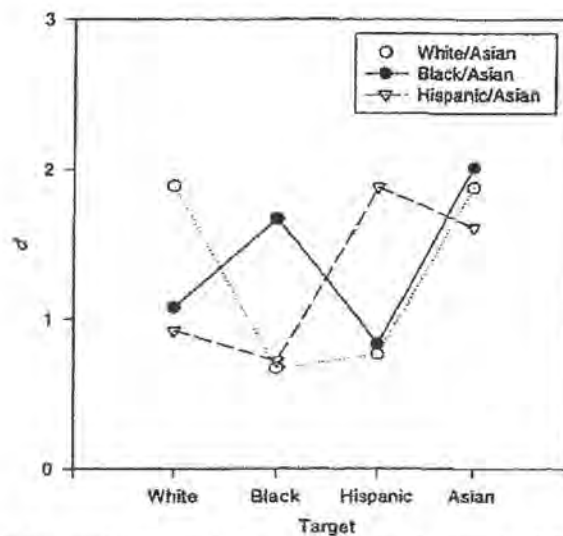|	Race of Face			
	Black		White	
Race of Observer	PL	UNPL	PL	UNPL
Black	1.21	1.48	0.96	1.13
White	0.98	1.09	1.25	1.39
Latino	0.92	1.10	1.30	1.46
Asian	0.93	1.02	1.16	1.31

NOTE: PL = pleasant mood, UNPL = unpleasant mood.

Anderson (1999) tested thirty White, Black, Hispanic, and Asian observers (presumably from New York as the research was conducted at Brooklyn College--N=120), with uniracial parentage (both parents of the same race). All observers were required to have lived in a uniracial home for the first three years of life. Each group displayed fairly similar impairment on cross-race faces (as reflected in a measure of recognition performance (d').
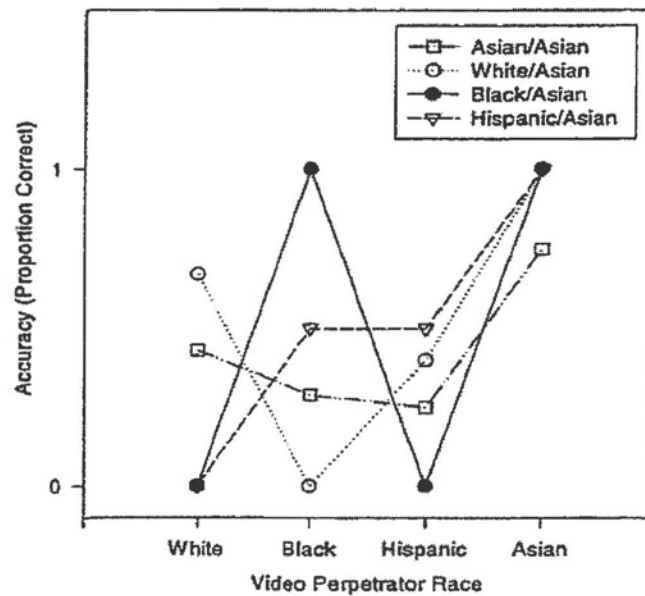
Participants with mixed parental backgrounds (N=107) were also tested and displayed impairment for groups not represented by one or the other of their parents. In particular, among Asian participants, the following figure shows that only participants with Asian/Black parent combinations did not show impairment on Black faces.

Asian Parent Combinations

Anderson also tested 227 participants (with varied parental racial mixes) ability to recognize the perpetrator (White, Black, Asian or Hispanic) in a videotaped robbery (in which the presence of a weapon was implied). The video lasted three minutes and 15 seconds during which the perpetrator was visible of two minutes and thirty seconds. Participants were tested with a six-person perpetrator-present photoarray. As the following figure illustrates, the same pattern of results emerged with respect to Asian participants and Black perpetrators: only participants with one Black parent outperformed other combinations.

Asian Parent Combinations

Gross (2009) also tested college-age participants from four major ethnic groups (White, Black, East Asian, and Hispanic) who were asked to view, remember, and later recognize faces from each of those groups. On average all groups had lived in California 11 years or more. As the following table demonstrates, there was significantly better performance (indexed by d'—with higher numbers indicating better performance) in all groups on faces from their own ethnic backgrounds—with particularly poor Asian on Black performance. As the second table shows, Asians were 50% more likely to misidentify a new Black face than a new Asian face.

Mean d'_e Scores as a Function of Participant Ethnicity
and Face Ethnicity

| Participant Ethnicity | Asian | Face Ethnicity | | |
		Black	Hispanic	White
Asian	.74 (.78)a	.18 (.63)b	.52(.77)c	.55 (.79)d
Black	.32 (.68)e	.80 (.84)f	.50 (.73)g	.85 (.90)h
Hispanic	.32 (.78)i	.12(.83)j	.78(.76)k	.64 (.81)l
White	.38 (.77)m	.33 (.79)n	.65 (.68)o	.86 (.81)p

Note. The following differences are significant at $p < .05$.
b < a, c, d; e, g < f, h; j < i < k, l; m, n < o < p.
e, i, m < a; b, j, n < f; j < n; e, g < k; d, l < h, p.
The following differences are significant at $p < .06$: c, d < a.

44

Proportion of "Yes," "Maybe," and "No" Responses to Old and New
Items as a Function of Participant Ethnicity and Face Ethnicity

Participant Ethnicity	Face Ethnicity	Old Items			New Items		
		"Yes"	"Maybe"	"No"	"Yes"	"Maybe"	"No"
Asian	Asian	.54	.17	.29	.20	.16	.64
	Black	.35	.24	.41	.30	.22	.48
	Hispanic	.42	.23	.35	.23	.15	.62
	White	.53	.16	.31	.24	.20	.56

Meissner & Brigham (2001) have recently written what I consider to be the best review of research on the problems of what have sometimes been called other-race or cross-race identifications or own-race bias (ORB). Meissner and Brigham analyzed data from 39 research articles, with 91 independent samples involving nearly 5,000 witness participants. Measures of correct identifications and false alarm rates, and aggregate measures of discrimination accuracy and response criterion were examined. Overall, they reported that when the perpetrator is present the ratio of correct to incorrect identifications was 40% higher for same-race identifications. The ratio of mistaken identifications to correct rejections in target-absent arrays was 56% greater for other-race identifications. Overall, the ratio of correct to incorrect identifications was more than 2.2 greater for own-race faces as compared with performance on other-race faces.

Meissner and Brigham reported that cross-race impairment was greater with short exposure times (especially with respect to false identifications), longer periods of time between viewing and identification were associated with larger ORB (especially with respect to false identifications), Whites showed more ORB than other groups (especially with respect to false identifications of Blacks).

Meissner and Brigham also explored the question of whether cross-race contact might reduce these effects and found that overall, contact played a small role in the ORB, accounting for just 2% of the variability across participants. They also found that the amount of viewing time available to witnesses significantly influenced accuracy in the ORB--particularly through an increase in false alarms to other-race faces when study time is limited. That is to say, there is little reason to think other-race contact would play an important role in reducing own-race bias and some reason to think the brief exposure to the perpetrator would aggravate the problem.

Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues reported that 51% of their respondents understood that cross-race identifications are less reliable than same-race identifications. As with the other factors discussed above, this means there is significant disagreement among laypersons about the impact that these factors have on witness confidence—eyewitness expert testimony would provide jurors with scientifically-grounded guidance on this issue.

Changes in Appearance/Disguise/Obscured Views

As noted at the beginning of Section VII above, the record indicates that the perpetrators were described as wearing hoodies and skull caps (hoodies verified by surveillance video). One witness noted the hoodies were pulled down and obscured the faces of the perpetrators.

Davis & Valentine (2009) observed: "There is extensive evidence that the removal or addition of disguises impairs facial recognition (e.g. wigs, hats, glasses: Diamond & Carey, 1977; dark glasses: Hockley, Hemsworth, & Consoli, 1999; Metzger, 2001; clear glasses and beards: Terry, 1994; for a meta-analysis see Shapiro & Penrod, 1986). In addition, Henderson et al. (2001; Experiment 3) found that unfamiliar simultaneous face matching accuracy was worse when actors were disguised with hats." (p. 491)

Changes in Appearance. In Shapiro and Penrod's (1986) meta-analysis, experiments were coded for whether or not the facial stimuli had undergone changes in facial features between the encoding and recognition phases. Facial transformations included changes in facial hair and deliberate disguises such as those used in the experiment just described. Nontransformed faces were more accurately recognized (d = 1.05; 75% non-transformed vs. 54% for transformed) and less often falsely identified (d = .40; 22% vs. 30%) than transformed faces.

A number of studies have shown that changes to the hair have a negative impact on recognition (Memon & Gabbert, 2003; Peterson & Baddeley, 1977; Pozzulo & Balfour, 2006; Pozzulo & Marciniak, 2006; Terry, 1994). Studies have similarly found differences in recognition accuracy for individuals wearing eyeglasses as compared to those not wearing eyeglasses (McKelvie, 1988; Metzger & Bridges, 2004; Patterson & Baddeley, 1977; Terry, 1993). Terry (1994) found that the removal of eyeglasses had a more negative effect on recognition accuracy than did their addition. Patterson & Baddeley (1977) found that the addition or removal of a beard (including mustache) substantially reduced recognition accuracy. Terry (1994) reported that the addition of a beard had a greater negative effect on accuracy than the removal of a beard.

Concealment. Full face masks stockings, hats and hoods can be quite effective in diminishing the facial feature cues that are necessary for recognition. In my research with Cutler (Cutler, Penrod, & Martens, 1987a, 1987b; Cutler et al., 1986; O'Rourke et al., 1989) we examined the effects of masking a target's hair and hairline cues on subsequent identification accuracy. In these experiments participants viewed a videotaped liquor store robbery and later attempted an identification from a videotaped lineup. In half of the robberies the robber wore a knit pullover cap that covered his hair and hairline. In the other half the robber did not wear a hat. The robber was less accurately identified when he was disguised. For example, in one of the experiments (Cutler et al., 1987a) 45% of the participants gave correct judgments on the lineup test if the robber wore no hat during the robbery, but only 27% gave a correct judgment if the robber wore the hat during the robbery. Pozzulo & Marciniak (2006) reported that correct identifications fell by half when a target's hairstyle changed (short versus long hair).

Metzger (2001) found the addition of sunglasses to composite faces significantly reduced recognition accuracy. Hockley, Hemswoth and Consoli (1999) found that sunglasses reduced correct identification rates and increased false identifications. Mansour et al. (2012) found that lineup identification accuracy was reduced when perpetrators wore sunglasses—20% fewer correct identifications and 6% fewer correct rejections.

Following work by Davies & Flin (1984), Mansour, et al., (2012) also examined the effects of stockings pulled over the head (and the effects of a variety of other disguises--as shown in the figure below which is taken from Figure 2 of the Mansour, et al., 2012 article). As the figure shows, the stocking they used blurred the facial characteristics of the target individual and compared to no-stocking views, correct identifications dropped from 80% to 55% and false identifications went from 21% to 36%.

More extreme disguises can produce even more substantial impairment. This is illustrated in a study Fisher and Cox (1995) where celebrities, known to the research participants, were identified by only 12% of participants when shown only the celebrities' eyes (mouths alone produced recognition by 3% of participants). Eyes+noses+mouths yielded 57% recognition (versus 100% when full faces were shown). The figures below provide a sense of the increasing levels of exposure which produce increasing levels of accuracy. Of course it is important to emphasize that the participants were endeavoring to identify persons known to them through multiple exposures over extended periods of time—not the faces of strangers seen briefly.
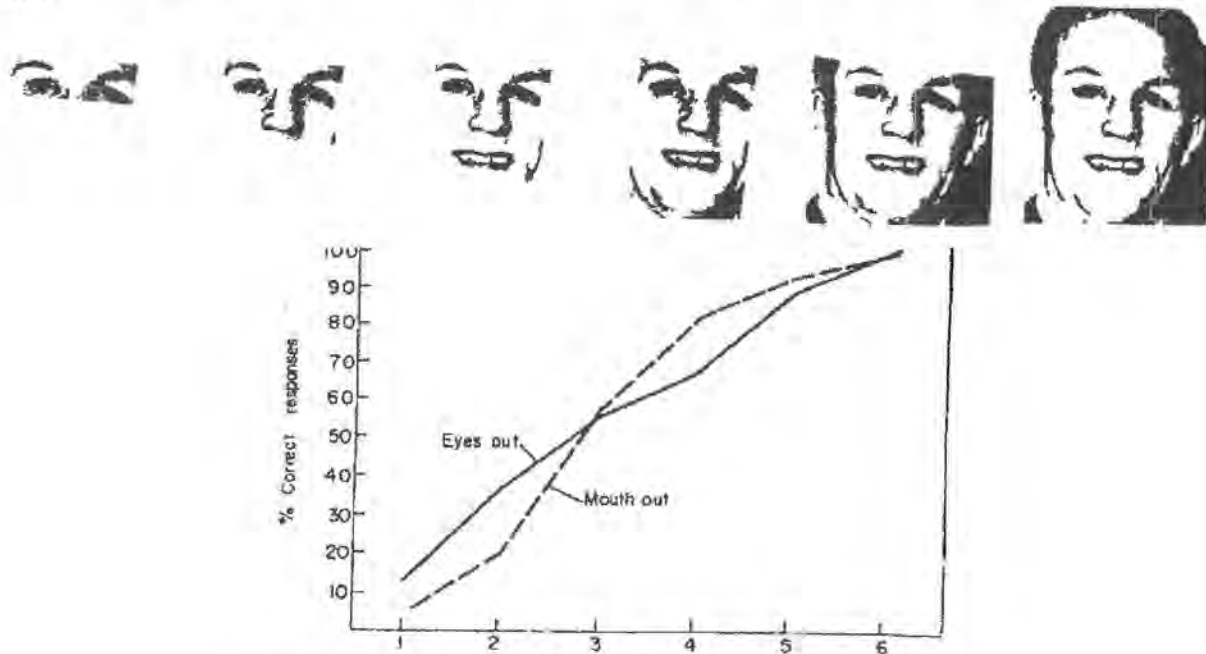

Fig.3   Stages of recognizing 'Eyes-Out' and 'Mouth-Out'
series of photographs

Neither Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues nor Benton et al. (2006) provides insight into lay understanding of the effects of disguise on identification accuracy. This does beg the question of whether laypersons would have any clue about the abominable performance shown in the Fisher and Cox study of celebrity recognition from eyes and mouths!

Retention Interval

As noted at the beginning of Section VII above, the record indicates that six to eight weeks passed between the time of crime and the identifications.

Common sense tells us that memory declines over time. But common sense does not tell us about the pattern of memory decline. Basic research in memory had long-ago documented the loss of memory over time (with a typical pattern being one of rapid early loss of information followed by a more gradual decline). Can we expect eyewitness identification accuracy to decline as the time between the crime and the identification test increases? My student Peter Shapiro and I included retention interval in our 1986 meta-analysis. When studies that manipulated retention interval were grouped into long versus short time delays, longer delays led to 10% fewer correct identifications and 8% more false identifications. Across all experimental cells in all

the studies examined in the meta-analysis (including those that did not directly manipulate retention interval) retention interval also proved an important determinant of correct identifications (r=-.11, p<.05).

Studies which test recognition accuracy at varying time intervals can tell us more than the simple comparisons in my meta-analysis or the field study (which lump observations into just two categories). Classic research by Egan, Pittner, and Goldstein (1977) found that the percentage of mock witnesses making false identifications increased significantly over time: 48% at 2 days, 62% at 21 days, and 93% at 56 days. The percentage of subjects making no errors declined from 45% (2 days) to 29% (21 days) to 7% (56 days).

As illustrated in the following table, memory loss is most rapid in the first hours and days following an event. Tollestrup, Turtle & Yuille (1994) studied suspect identification rates (suspects may, of course, be innocent) in actual robbery cases—the sharpest drop in suspect identification rates was reflected in the comparison of .5 days delay versus 3.6 days (an average drop of 8% per day, versus an average drop of less than .3% per day over the next 116 days).

Delay	Average	Sample size	Proportion (percent)
Robbery			
0–1	0.5	14	71.43
3–5	3.6	15	46.67
7–34	18.9	21	33.33
38–191	120.21	14	14.29

Another study of suspect identifications from police files (Behrman & Davey (2001), showed that 55% of the witnesses picked the suspect from a photo lineup if the identification attempt was within a 0–7-day delay period (a time period in which substantial loss has already occurred) and 45% did so after a delay of greater than 7-days.

My colleagues Deffenbacher, Bornstein, McGorty and I (2008) demonstrate that loss of eyewitness memory is most rapid early on…in the first minutes and hours. The forgetting curves/loss of memory which emerges from the studies in our meta-analysis shows that memory has typically declined by 15%-20% within 2 hours, with a further 4-5% drop in the next 10 hours. From that point forward the loss is less dramatic…though it continues—as is evident in the studies cited above. The following figure plots the loss of memory found in studies (from the meta-analysis) with varying retention intervals.

Loss of Eyewitness Memory Over One Month

48

The rapid early loss of memory is illustrated by the results from a study by Yarmey, Yarmey & Yarmey (1996) shown in the table below.

Table I. Mean Proportion Identification Scores by Retention Intervals

	Retention interval			
	Immediate	30 min	2 h	24 h
Photo showup				
Target present				
n	33	33	35	36
Hits	.70	.64	.54	.55
Incorrect rejections	.30	.36	.46	.45
Target absent				
n	34	34	38	36
Correct rejections	.82	.56	.42	.47
False identifications	.18	.44	.58	.53
Photo lineup				
Target present				
n	35	34	33	34
Hits	.49	.39	.36	.32
Incorrect rejections	.23	.26	.18	.21
Foil identifications	.28	.35	.46	.47
Target absent				
n	37	39	35	39
Correct rejections	.38	.28	.31	.29
False identifications	.16	.33	.14	.14
Foil identifications	.46	.39	.55	.57

Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues reported that just 53% of American respondents understood the "forgetting curve" reflecting rapid early loss of information from memory. As with other factors this means there is widespread disagreement among laypersons about the pattern of memory loss and the effect of the passage of time on identification performance—eyewitness expert testimony would provide jurors with scientifically-grounded guidance on this issue.

IX. Aspects of Identification Procedures that Affect Identification Accuracy

Desiree Scroggins indicated in trial testimony that after she made her tentative identification of Mr. Nolan on January 27, 2014 from an all-suspect 22-person array, she went to her mother's room, that her mother knew Mr. Nolan's name at that point and together they looked at Facebook pictures of Mr. Nolan. (pp. 152-153).

These facts are apparently communicated to Christopher Martinez and Sandra Martinez who also view Mr. Nolan prior to their identification of Mr. Nolan on February 10, 2014:

Q. And at some point, you tell Christopher that you had seen this person's Facebook page, right?
A. I showed, I told Christopher I identified the person who hit me and started holding us up.
Q. And showed him the Facebook?
A. Yes.
Q. And he said that's White Boy from my old neighborhood?
Q. He wouldn't do this?
A. Yes.

49

Q. He said that to you, right?

A. Yes.

Q. He said the White Boy I knew wouldn't have done this?

A. Yes.

Q. And at some point, Sandra Martinez comes to the house, or was she already there?

A. No. She was -- no. She came to the house probably like an hour later.

...

Q. And so you told Sandra the same thing and you showed her the Facebook page, is that right?

A. Yes

...

Q. And you and your mother were on Facebook together, right?

A. Yeah, me and my mom were the ones on Facebook together. Then when I had went over to Christopher, I showed Christopher the Facebook as well.

Q. And this is on December 27, the same day -- I'm sorry,   January 27, the same day that you first saw those initial 22 photos, isn't that correct?

A. Yes. (pp. 172-174)

In his trial testimony Christopher Martinez responded as follow about the Facebook viewings on cross-examination:

Q. And before that, you didn't think you knew the person who had committed the robbery, isn't that correct, sir?

A. I didn't, yeah.

Q. And she showed you a picture of somebody you knew from way back, many years before, somebody you knew as White Boy?   Yes or no?

A. Yes.

Q. And you told her I don't think White Boy would do something like this, isn't that right?

A. Yes, I did say that.

Q. But you didn't say to her, well, I know White Boy did it because I now remember that he did it?

A. I didn't connect, in my memory, to the situation that happened and the way he was in the room.

...

Q. And you said that when you looked at the photograph that Desiree showed you, whenever that was, but before this identification procedure, you said to her I don't think White Boy would have done something like this.   Is that correct?

A. I don't, it's not I don't think. I believed it at the time.

Q. You didn't believe that White Boy would do it?

A. Yeah, but then I had to really think about it. Yeah. (pp. 267-268)

Multiple Suspect Arrays

What is the relevance of the 22-person all-suspect array? It is noteworthy that the first instruction in the NIJ Guide is to have only one suspect per array. To do otherwise—as was apparently done in this case—completely defies the logic of having a multi-person array. Multi-person arrays are intended to protect innocent suspects from over-eager and poor-memory witnesses. If a witness makes a "guess" in a one-suspect six-person array where the suspect is innocent, there is only a one in six chance they will pick the suspect. If everyone in the array is a suspect (and in the worst-case scenario, they are all innocent) an innocent person will be selected every time a choice is made. The protections afforded to innocent suspects are dramatically diluted in an all-suspect array—it is akin to showing six showups all at once. The foulness of this practice is underscored by the research cited above which shows that nearly 40% of witnesses who make identifications from single-suspect arrays are indisputably mis-identifying innocent people – fortunately those innocent people are known to be innocent because they were selected to serve as innocent

foils. What do we imagine those 40% who pick innocent fillers are going to do if presented with all-suspect arrays!?!?

Prior Exposure Effects

We know from a variety of studies that prior viewings of an innocent person can render them vulnerable to a significantly enhanced risk of misidentification. One source of such effects is mugshot viewings about which Deffenbacher, Bornstein and I wrote in 2006:

The phenomenon of the photobiased lineup has existed for as long as police have, in the furtherance of their criminal investigations, engaged in the practice of exposing eyewitnesses to mugshots. If a witness is exposed to mugshots subsequent to viewing a perpetrator and prior to an additional test of recognition memory, there is the possibility that exposure to the mugshot photos may bias the witness's decision at that test, usually a photo or live lineup. Defense attorneys have labeled a photobiased lineup as one in which a possibly innocent person has been arrested because his/her photo was initially included in a set of mugshots but not identified by the witness; subsequently, however, he/she was identified by that witness at a lineup. Any bias thus engendered would be due to the previous exposure of the witness to the mugshot photo of the suspect identified at the subsequent lineup.

In such cases, the witness may well have suffered a failure of memory for the circumstances of previous encounter of the face identified at the lineup. This sort of failure of memory for facial source or context is all the more problematic when viewing of the perpetrator has occurred under less than optimal viewing conditions (Brown, Deffenbacher, & Sturgill, 1977). Indeed, the United States Supreme Court had opined that this was the case, even before there was empirical support for their view (*Simmons v. United States*, 1968).

The first empirical examination of the possibility of bias imposed by mugshot exposure more generally as well as the more specific sort of bias that might be engendered by a photobiased lineup was that of Brown et al. (1977). In all three of their experiments, these investigators noted that recognition memory for a once-seen face is much greater than is the capacity to recall the circumstances under which that previously unfamiliar face had been encountered. In Experiments 2 and 3, they employed what we will henceforth refer to as *a transference design* to test for specific biasing effects of mugshot presentation between exposure to live targets and a subsequent lineup. ...

At the lineup, eyewitnesses in Experiments 2 and 3 of Brown et al.'s (1977) study did indeed identify persons seen previously only in mugshots at significantly higher rates than they did previously unfamiliar foils. Interestingly, under Experiment 3's viewing conditions of only a brief glimpse of target persons, witnesses were as likely to indict a person on the basis of a single mugshot encounter as they were a target person encountered only once, that is a target person not also presented in mugshots.

...Deffenbacher, Carr, and Leu (1981) tested whether a transference design would produce significant loss of memory for context of facial encounter at retention intervals of 2 min and 2 weeks. At both retention intervals, they found significantly higher false alarm rates to faces seen once previously in mugshots than to previously unfamiliar foils. ...

In addition to the passive effect of familiarity gained by mere mugshot exposure to a previously unfamiliar face, Gorenstein and Ellsworth (1980) sought to determine whether there was a further increase in interference with witness memory for the target face engendered by the witness actively choosing the mugshot of an innocent person. Would witness commitment to a prior choice of a particular mugshot as the target person cause the mugshot image to be retained as strongly or even more strongly than that of the target? In brief, then, a mugshot commitment design involves a control group and an experimental group both being exposed to a target person, followed by mugshot exposure for the latter group. Witnesses in the experimental condition may be urged to pick an innocent person from the mugshots, despite the target's face not being included. At the lineup the target face may or may not be included, but the previously chosen mugshot image definitely is.

Gorenstein and Ellsworth (1980) found that whereas control witnesses identified the target from a six-person photo array at a significantly above chance rate (.39), mugshot commitment witnesses failed to identify her at an above chance rate (.22) but did identify their mugshot choice's image at an above chance level (.44), a level somewhat greater than the rate at which control witnesses identified the target person. (pp. 288-290).

In our (2006) meta-analysis of 15 mugshot studies Deffenbacher, Bornstein and I found the overall proportion of correct identifications from target-present and target-absent arrays following mugshot exposure was .43 whereas the comparable figure for the no-mugshot control conditions was .50 correct--indicating that mugshot exposure inserted between exposure to the target and a subsequent lineup test has a negative effect on eyewitness memory. In 11 of those 15 studies only innocent people were shown in the mugshots and the negative effects of exposure were even greater: the proportion of correct identifications from target-present and target-absent arrays following mugshot exposure was .48 versus .61 for the no-mugshot control condition. If the witness actually made an identification of a mugshot image as that of the target, the subsequent appearance of that image or person at an identification procedure had an even larger negative impact upon lineup accuracy, with mean proportion correct of .32 for the mugshot commitment condition versus .55 for the control condition. Identification of an innocent mugshot interferes with a witness's memory of the actual target so as to reduce both the hit rate and correct rejection rate for target present and target absent lineups, respectively.

We also specifically examined witness performance when they are shown mugshots and a subsequent presentation that does not contain the target/perpetrator. The overall false alarm rate at a lineup for mugshot exposure conditions was .37, and the comparable figure for the no-mugshot control conditions was .15, less than half the former rate. More alarmingly, if the witness selected an innocent person in a mugshot and that person was included in a subsequent target-absent array, the mean false alarm rate for the mugshot commitment condition was .53 versus just .20 for the control condition.

Goodsell, et al. (2015) further illustrates the problem—in their study 85% of participants made an erroneous choice from a 50-photo mugbook following a 25 second view of a perpetrator and the question was what those witnesses would do when shown a photoarray containing there erroneous choice or a replacement, the perpetrator or another look-alike, a non-selected filler from the mugbook and several new faces: "A commitment error occurs when participants search through mug shots for the perpetrator, make a selection from the mug book, and then select that same individual from a subsequent lineup, even when faced with the actual perpetrator. The present study replicated a performance decrement arising from the commitment effect. ...Commitment resulted in participants not recognizing the perpetrator, which supported a memory-blending/replacement hypothesis." (p. 219)

As the following table shows—among witnesses shown an array including their erroneous choice (the CI column), nearly 70% stayed with their erroneous choice (vs 8% who shifted to the perpetrator and 11% who picked a different foil). When the perpetrator was not in the array, 81% stayed with their erroneous prior choice and 14% picked another innocent foil. When the erroneous choice was not included in the array (CNI column) there were more perpetrator choices (18%) but even more foil choices (the previously vied but nont-selcted foil jumped to 28% and other foils drew 15% of choices). When neither the previously chosen foil nor the perpetrator was present, 38% picked a foil they had seen in the mugshots. Goodsell, et al. conclude: "This study provides additional evidence that once an eyewitness has made an identification, no further identifications should be attempted because access to memory for the perpetrator is diminished." (p. 219) ... "Our results are consistent with prior research on mug shot exposure and suggest that witnesses who are exposed to a mug shot search should not participate in a subsequent lineup identification task. Witnesses who commit to a prior choice are unreliable because their memory for the perpetrator is affected. Those who made a commitment error failed to recognize the actual perpetrator..." (p. 231).

Table 2. Proportion of lineup choices by condition.

		Perpetrator search	
	Control	CI	CNI
Perpetrator present			
Perpetrator	.294	.083	.175
Selected foil	n/a	.695	n/a
Familiar foil	n/a	.028	.275
New foil	.588	.083	.150
No ID	.118	.111	.400
N	51	36	40
Perpetrator absent			
Selected foil	n/a	.814	n/a
Familiar foil	n/a	.000	.378
New foil	.86	.139	.178
No ID	.14	.047	.444
N	50	43	45

CI = choice included, CNI = choice not included, n/a = not applicable.

Desmarais & Read's (2010) meta-analysis of lay respondents' knowledge of eyewitness issues reported that 61% of American respondents understood mugshot exposure can increase the risk of misidentifications. This means there is still substantial disagreement among laypersons about the negative impact that mugshot exposure can have on identification performance—eyewitness expert testimony would provide jurors with scientifically-grounded guidance on this issue and, as detailed above, increase juror sensitivity to its effect.

(Mis-)Recognition of "Familiar" Faces

As noted above, Christopher Martinez and Sandra Martinez indicated, on February 10, 2014 when they identified Mr. Nolan that they had known him some years earlier from the neighborhood but had not seen him in recent years. The witnesses did not recognize either of the perpetrators as Mr. Nolan at the time of the home invasion nearly two months earlier (12/16/13).

Are judgments that a face is familiar reliable? Factors other than actual familiarity can influence the perception of familiarity and (mis)-recognition. For example, Vokey & Read (1992) demonstrated that typical-looking faces were associated with stronger perceptions of general familiarity and led to false recognitions. They report: "for every unit increase in general familiarity, the false alarm rate increased by 5% or 6%" (familiarity was measured on 4-point scales, so the effect could be quite large).

The latest relevant study is by Pezdek & Stolzenberg (2014). Their results confirm the earlier findings and are particularly relevant to the present case--here are excerpts from the abstract--- *sophomores (N 139) in two small private high schools viewed yearbook pictures of (a) graduated students from their school who were seniors (fourth year) when participants were freshmen (first year) (familiar) and (b) unfamiliar individuals, and responded whether each was 'familiar'. The design was completely crossed; familiar faces at each school served as unfamiliar faces at the other school. Although individuals' familiarity judgments were diagnostic of prior contact, accuracy was low (mean hit rate 0.42; mean false alarm rate 0.23), rendering an eyewitness's report of having seen a perpetrator casually in the past of limited forensic value.*

Research shows that, counter-intuitively, identifications of "familiar" faces - as with identifications of unfamiliar faces - are influenced by a host of variables, including interaction time, contextual information, expectation, postevent information, and own-race bias. ... these variables simultaneously increase the

probability that an individual will identify a given face as familiar and inflate the individual's confidence in the identification, regardless of whether the face in fact is familiar or the identification accurate.

From the outset it is important to understand that witnesses in cases where the witness claims to know the perpetrator regularly make identification errors. In a 2003 study of 640 actual eyewitness identification attempts—56 of which involved suspects putatively known to the witness--Valentine et al. found that in 21% of cases where the suspect was supposedly known, the eyewitness failed to identify the suspect and in another 5% of cases the witness identified an innocent filler. In a larger study with 632 witnesses who said they knew the perpetrator (Memon et al, 2011) a comparable percentage--6.1%--of those witnesses misidentified an innocent filler.

Factors known to influence stranger identifications can similarly influence non-stranger identifications. Kersholt et al. (1992) report: "For commonplace situational factors like poor viewing conditions and facial typicality, what is surprising is the magnitude of the deleterious effects. In one study, bad lighting and partially concealing the subject's face and hair reduced correct identifications of known subjects by 18 percent."

These results raise the question: why do witnesses misidentify people who are supposedly known to them?

A 1985 study by Young et al. (1985) included 922 records of instances in which people had difficulty recognizing other people. These included 314 misidentifications--272 instances in which an unfamiliar person was misidentified as a familiar person and 42 mistaken identifications of one familiar person for another. Young et al.'s summary (p. 505) of these cases is informative both with respect to the circumstances that gave rise to the errors (akin to the conditions in the present case) and the circumstances which gave rise to corrections of errors (*not* akin to the present case) : "...the diarists often tended to think that the unfamiliar person they misidentified was someone who was highly familiar (54 per cent) and often seen (more than once a week in 45 per cent of the records). ... In short, then, the most salient features of misidentifications of an unfamiliar person as a familiar person are that they tended to be brief, and were often associated with poor conditions; they usually ended when a better view was obtained."

It turns out that prior familiarity can actually *reduce* the reliability of an eyewitness identification. A 1995 study by Read demonstrates that greater familiarity can sometimes lead to greater error. Read's study was an "experiment that examined the effect of increased interaction time on store clerks' ability to accurately recognize and identify a female target. Half of the clerks spoke with the female target for 30-60 seconds ("short-interaction" clerks); the other half were interviewed by the female target for between four and 12 minutes ("long-interaction" clerks). Two days later, all clerks were asked to pick out the targets' photographs in various photographic lineups; some lineups contained the targets' photos (target-present), while others did not (target-absent). On the whole, the most reliable effect of increased exposure to the female target was the clerks' greater willingness to identify someone from the lineups, but their choices were wholly unrelated to increased ability to choose the correct photograph."For both target-present and target-absent lineups, the long-interaction clerks more often erroneously picked the photograph of a stranger as the interviewer. Moreover, the long-interaction clerks' performance deteriorated most severely – and most disturbingly for eyewitness identification purposes - in target-absent lineups: the percentage of clerks who incorrectly said the target was present more than doubled as their interaction time with the female target increased." (Coleman, et al. p. 53).

The misidentification error rates in the Read study were not trivial—running as high as 50%.

What the jury needs to hear in the present case is how witnesses can be wrong in what are asserted to be non-stranger cases, so they can judge the reliability of the identification in the present case instead of assuming that the witness is correct because the witness asserts they are familiar with the defendant. Part oft he point of expert testimony in the present case would be to explain factors which have been shown to

influence stranger identification accuracy and which, on logical grounds—like lighting and concealment—could influence non-stranger identifications.

Unconscious Transference

Face familiarity can also induce what has been referred to as "unconscious transference" –a term applied when there is a "transfer of one person's identity to that of another person from a different setting, time, or context" (Read et al., 1990, p. 3). Williams (1963) originally coined the phrase in reference an English murder case, in which an innocent person may have been executed. Williams indicates that one of the eyewitnesses who identified the defendant had briefly seen him before the crime and may have unconsciously transferred the identity of the actual perpetrator to the defendant.

Bystander and Prior Viewing Errors. Studies analogous to the case Williams (1963) described have been conducted by a number of researchers. A classic study of the confusion of roles was conducted by Buckhout, et al. (1974) who staged a mock assault before a group of 141 students. When the students were later asked to pick the assailant from a group of six photographs including the assailant, an innocent bystander to the assault and four fillers, 40% were able to select the assailant and 25% selected the person who had been at the scene as an innocent. Loftus (1976) demonstrated a substantial transference effect in which a bystander was selected at greater than chance rates by witnesses from a target-absent array. In that study college students listened to an audiotape of a crime and photographs from a high school yearbook illustrated the characters in the story. After three days participants attempted to identify the assailant from one of two lineups which contained the assailant plus four unfamiliar faces, the face of an innocent bystander who was seen in the story plus four unfamiliar faces. When the assailant was in the lineup, 84 percent of the participants correctly identified him. However when the bystander was in the lineup, 76 percent of the participants misidentified someone, and 60 percent of these participants misidentified the familiar bystander.

Read et al. (1990) conducted five field experiments involving the prior viewing of witnessing of bystanders and targets by retail store clerks or by students in university classrooms going about their normal situation-appropriate activities. Tests of the difference between transference and control conditions in regard to the proportion of false positive identifications of the bystander produced evidence of transference effects. In one of their experiments control and transference participants viewed an "assailant" (an electrician who interrupted class to fix an electrical problem) and transference participants were also exposed to a bystander who was seen distributing exam materials in another classroom); control participants did not see the bystander. Two weeks later participants attempted to identify the assailant from a photoarray containing the bystander and four unfamiliar foils. Twenty-five percent of the transference witnesses misidentified the bystander, as compared with just 12 percent of the control subjects. In addition, 13 percent of the transference participants who misidentified him thought the assailant and the bystander were the same person and they had seen him in two different places.

Ross, Ceci, Dunning and Toglia (1994) students watched a film with a series of short segments showing teachers interacting with children. At the conclusion of the film a female teacher walks into a cafeteria and is robbed. Everyone saw the same film but transference witnesses also saw a male bystander to a group of children several minutes before the robbery. The assailant and bystander were both seen for about thirty-four seconds. Between three to five minutes later participants were asked to identify the assailant from an array that contained the bystander and four unfamiliar foils. Transference witnesses were nearly three times more likely to misidentify the bystander than control participants, (61 percent versus 22 percent) and were less likely to respond correctly that the assailant was not in the lineup (34 percent versus 64 percent). Sixty-six percent of transference participants mistakenly believed they had seen the assailant outside the cafeteria -- 95 percent mistakenly said the assailant was reading to the children. Few of the control participants (4 percent) made this error.